Models of Moral Cognition

Jeffrey White

Abstract This paper is about modeling morality, with a proposal as to the best way to do it. There is the small problem, however, in continuing disagreements over what morality actually is, and so what is worth modeling. This paper resolves this problem around an understanding of the purpose of a moral model, and from this purpose approaches the best way to model morality.

1 Introduction

The process here analyzed is not a dream, a fancy floating in the air; it is perfectly real, and by no means infrequent.

—Schopenhauer[1]

A model is a representation of salient aspects of a system that, when rendered together, articulate an essential function in a more efficient way than the original, a replica or a duplicate. So, models are created for reasons other than for the creation of one of these other things. Some models are explanations. For example, a model of disease represents how pathology progresses. Some models are made to help realize an original. For example, models of buildings inform architects and engineers how to make original buildings which, once constructed, can serve in the creation of duplicates or replicas. Models of this sort are especially important when new answers are necessary, novel creations in response to new problems and the questions that these raise. This paper is interested in models that do this sort of

[1] [1], p. 170.

J. White (✉)
KAIST, Daejeon, South Korea
e-mail: jeffreywhitephd@gmail.com

L. Magnani (ed.), *Model-Based Reasoning in Science and Technology*,
Studies in Applied Philosophy, Epistemology and Rational Ethics 8,
DOI: 10.1007/978-3-642-37428-9_20, © Springer-Verlag Berlin Heidelberg 2013

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 20 | Date: 9-5-2013 | Page: 2/28

work, but rather than help in building better houses, the models that we are after should help us to become better people. Rather than model new places to stay, new futures to grow into, ourselves included.

Two general forms of moral model are prevalent, and both seem to aid moral development. The traditional form is one of narrative and ethical theory expressing principles affirmed by intuition and enculturation through example, demonstration, and argument, and the other, more recently popular form is that of mechanistic and information processing models of specific subroutines and circuits within the brain, within the organism, or within the extant ecosystem, all working together to tell the story of morality. Which mode of representation is best?

Twenty years ago, anticipating the impact of the cognitive sciences on moral philosophy, Stephen Stich asked a similar question, and pointed in the direction of psychological representations. A quick review of his reasons for this will help to provide some context for the rest of this paper, as well as set up some important issues to be met with along the way, including the role of models in moral practice, and potential for future research.

2 Looking for Mr. Goodmodule?

In a talk from 1989 published in 1993, Stephen Stich argued that a central project in traditional moral philosophy had been chasing its tail, and issued a sort of rallying call to future-minded moral philosophers around a forecast that "the beginnings of moral philosophy fall squarely within the domain of cognitive science"[2] [2]. Stich argued that moral philosophy had been searching after things that "do not exist," and he identified a set of "Platonic assumptions" responsible for leading the inquiry astray. The first problem was that some philosophers had "presumed that the mental structures underlying moral judgments are rather like definitions" in that they "specify individually necessary and jointly sufficient conditions for the application of moral concepts." The second problem was the claim to reliable intuitions about these definitions, with the "central strategy in testing a proposed definition" being merely "to compare what the definition says to what we would say about a variety of actual and hypothetical cases." And, the third assumption that Stich found active was mistaking the central task of moral philosophy to be "making explicit the necessary and sufficient conditions that, presumably, we already tacitly know," (pp. 3–4) Thus, we see moral philosophy setting out for itself both the terms of its own inquiry and the standards for their evaluation. Self, chasing, tail.[3]

[2] [2], p. 14. Noted pagination belongs to the author's copy, a copy of which is maintained by Joshua Knobe online at the address cited.

[3] It is as I had read the other day, "It is a familiar problem in recent philosophy that to the extent my experience of another person can be assimilated to ready-made experiential categories, I have

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 20 | Date: 9-5-2013 | Page: 3/28

Rather, Stich saw the future of moral theory in psychological alternatives "that do not involve necessary and sufficient conditions." These aim to represent moral concepts in forms that people already comfortably employ in directing and evaluating everyday morally *in*significant action, like "the knowledge structures that guide our expectations in reading stories about restaurants and other common social situations."[4]

Stich found that these everyday frames, as well as other systems of representation under psychological consideration—"Mathematical knowledge, knowledge of various sciences, and common sense knowledge in various domains" (p. 13)—are analogous to moral systems in a very important way, in that people

> *can offer a complex, subtle, and apparently systematic array of judgments about particular cases, with little or no conscious access to the mechanisms or principles underlying these judgments* (pp. 13–14).

This fact sheds some light on the purpose of moral philosophy, as well as on the structure of moral judgment. Moral judgment is the product of something deeper, that informs consciousness. And, the best ways to represent morality are those ways that best communicate the significance of these deeper things. Stich approached these issues through his primary vocation, as an ethics teacher. Given the purpose to effectively communicate moral concepts, truths, so that students can assess, assimilate, and critically evaluate morally salient situations, thereby empowered through understanding to a lifetime of free philosophical self-determination, the best way to represent morality is easily determined. In the same ways that people demonstrate, learn and understand morality, already, through direct and indirect experience of the moral lives of self and others:

> *Exemplar models of conceptual representation, and more sophisticated variations on the theme that invoke "scripts" or stories, also suggest an explanation for the fact that those engaged in moral pedagogy generally prefer examples to explicit principles or definitions. Myths, parables, fables, snippets of biography (real or fanciful)—these seem to be the principal tools of a successful moral teacher. Perhaps this is because moral knowledge is stored in the form of examples and stories. It may well be that moral doctrines cast in the form of necessary and sufficient conditions are didactically ineffective because they are presented in a form that the mind cannot readily use* (p. 11).

An exemplar is a "specific instance" of some unique thing "falling under a concept." On the view that concepts are represented in the form of exemplars, "categorization" of perceived objects

> *proceeds by activating the mental representations of one or more exemplars for the concept at hand, and then assessing the similarity between the exemplars and the item to be categorized* (p. 10).

(Footnote 3 continued)
not really gotten beyond myself. Rather, in the experience of the apparent other, I have merely reconfirmed or reconnected with a prior sense of self-identity." [3], p. 119.

[4] Today, this is deeply researched, with some technology employed in reading minds by reproducing field effects within areas of the brain matching those of the donor.

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 20 | Date: 9-5-2013 | Page: 4/28

In this way, exemplars serve as vehicles for moral knowledge by demonstrating modes of being through which moral concepts are expressed. But more than that, exemplars are a special kind of model, for they are something that one can "model" himself after. One can direct one's life along similar paths as those demonstrated by exemplars, mimicking routine actions in routine contexts, and one can compare one's self against exemplified demonstrations as standards. In the comparison, one feels what it is like to differ from these examples, feeling the difference as the satisfaction or failure to meet exemplified standards. How well exemplars work applied to novel situations, however, is another problem, and one that we will come to as this paper closes.

The space of academic philosophy offers the leisurely reflection necessary for ready analysis of possible situations and application of principle, where exemplars are not such efficient vehicles of knowledge.[5] This is clearly not the most efficient way to model morality, however, unless expecting everyone with moral aspirations to spend their days engaged in professional moral philosophy. Rather, exemplar models work in everyday life because moral knowledge is about human lives, and human life is more effectively represented in examples and demonstrations than categories and principles.

Since the time of this writing, great progress has been made toward confirming Stich's forecast that "the beginnings of moral philosophy fall squarely within the domain of cognitive science." By 2000, Nancy Eisenberg was able to report that "Philosophers' changing view of the role of emotion in morality is consistent with the predominant view of emotion in psychology today" in understanding that "higher-order emotions such as guilt and sympathy are believed to motivate moral behavior and to play a role in its development and in moral character" ([4], p. 666). And, since 2000, the area between moral philosophy and the cognitive sciences has exploded, with disciplines at its core notably absent from Stich's short list of philosophy, anthropology, and cognitive psychology. To these must be added a cluster of new fields falling directly under his forward gaze, experimental philosophy, neurolaw, neuroethics, neurophenomenology, and social cognitive neuroscience amongst them, all with a focus on correlating "what it feels like" with neural activity understood either metabolically or computationally. All of this confirming Stich's:

> *strong suspicion that progress in understanding how people represent and use moral systems will not be made until scientists and scholars from these various disciplines begin to address the problem collaboratively. Indeed, one of my goals in writing this chapter is to convince at least some of my readers that it is time to launch such a collaborative effort* ([2], p. 14).

Here is a short list of traditional philosophical terms that are being naturalized through ongoing interdisciplinary work around the issue of moral cognition:

[5] And, as aging studies have shown, older people tend to rest on old ideas, with aging lazy philosophers, hashing out the fine points of established definitions is expected according to brain research.

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 20 | Date: 9-5-2013 | Page: 5/28

- "experience" as "conditions under which associations are formed between novel stimuli and biologically innately significant events, typically innate triggers," ([5], p. 656)
- "intuition" as product of one thread of the dual-processing portrait, "associative" and "attuned to encoding and processing statistical regularities, frequencies, and correlations in the environment," ([6], p. 990)
- "moral intuition" as "fast, automatic, and (usually) affect-laden processes in which an evaluative feeling of good-bad or like-dislike (about the actions or character of a person) appears in consciousness without any awareness of having gone through steps of search, weighing evidence, or inferring a conclusion," ([7], p. 998)
  - with the "key functional difference" between moral and other intuitions being "that moral intuitions appear to make a difference, directly, to how we act and react," ([8], p. 7)
- "moral emotion" as an extension of root-level survival circuits distributed throughout the body and realized in the brain as emotions that are at once evaluating and motivational, [5]
  - with Jonathan Haidt confining the moral to just those emotions that are concerned with others rather than with one's own prudential self [9].
- and most recently "conscience," "a neural process that generates emotional intuitions combining somatic perception (the gut reaction) with cognitive appraisal concerning a special subset of goals"([10], p. 156).

When Stich was writing, without models of neural processing assembled from basic neurological research, it had been easier to conceive of universally binding rational principles than similarly effective sets of somato-affective markers and their corresponding motivations. Traditionally, intuitionist, sentimentalist, or emotivist accounts of moral cognition had been hamstrung by a limiting capacity to draw their subjects in clear and distinct terms. Now, the "new synthesis" in neuroethics promises to open new avenues to toleration, compassion, and mutual understanding built on what is best understood as the "shared body."

Not confined to individual human agency, neurological research has also informed thinking on the issue of collective agents, where mirror systems and empathy embodied in individual subjects help to explain inter-subjective associations whereby "Some people may act "as-if" a certain belief was their own without actually endorsing it themselves," with the result being the appearance of unity, and so of collective agents as entities in their own right.[6] Thusly, through advances in functional imaging, a real-time picture of man's moral reality built of affect, bottom-up, is being extended from neural substrate to intuition to institution

[6] [11], p. 336. Such tendency to social coherence is also affirmed in the cognitive "switch" that turns individual fans into a seething mass, helping to explain the loss of self also experienced by persons caught up in the mass psychology of crowds.

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 20 | Date: 9-5-2013 | Page: 6/28

and social organization, deep in territory traditionally belonging to moral philosophy.

In this spirit, Young and Saxe point out that individual differences in moral judgment can be mapped onto regularly recurring patterns and intensities of activity in different areas of the respective subjects' brains. These differences correlate with education, upbringing, and routine attitude, and even characteristic mood, with Saxe reminding us, for example, that "people who are generally disgusted make harsher moral judgments of unrelated incidents." Their approach is to discover such patterns of activation common between individuals and groups, thereby revealing the "independent psychological components of moral judgments" and the neurological basis for "apparently arbitrary "cultural clusters" of moral value." Ultimately, Saxe, suggests that mapping neural differences between parties to moral differences, "may help us to understand and resolve moral disagreements not only between individuals but also on a broader scale." She, as Stich two decades prior, points to the future of moral inquiry in psychological representations, forecasting that "The next stage for research must therefore be to understand the structures underlying these differences" ([12], p. 324).

Pursuit of the mechanisms underlying moral judgments may reveal a universal basis for moral judgment in these same mechanisms, with the hope that these provide all that is necessary for moral guidance. Consider Jonathan Haidt's assessment of the relative importance of intuition and moral reason in that effort:

> *In other words, evolution shaped human brains to have structures that enable us to experience moral emotions, these emotional reactions provide the basis for intuitions about right and wrong, and we* (*or, at least, many moral theorists*) *make up grand theories afterward to justify our intuitions* ([9], p. 68).

And, Cokely and Feltz second this sentiment, suggesting that not only are these theories post hoc, but they may also be counterproductive:

> *In an uncertain and complex world such as ours, we should not expect or necessarily even want to always be governed by processes that maintain logically coherent cognition* ([13], p. 358).

This is a long way from Stephen Stich rejecting necessary and sufficient conditions as necessary and sufficient for moral theory. In the words of Darcia Naevaez, "the pendulum is swinging in the other direction and reasoning is often considered unnecessary" ([14], p. 164).

It may be that remaining rational is not always rational. And, understanding the grounds for moral differences through somato-affective mechanisms is a long way from the high point of the rationalist pendulum in the other direction. But does distance equate with progress? Rather, it is my strong suspicion that progress on the issue of moral representations cannot be made unless the highpoints of either are reconciled with one another, seconded by an even stronger suspicion that we have been in this situation, before.

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 20 | Date: 9-5-2013 | Page: 7/28

As an example of a rationalist high point in moral theory, consider the following conclusion from Hastings Rashdall on the plausibility of intuitionist theories that morality is an emotion:

> *I have tried to suggest to you that they can be met in as purely a scientific and dispassionate manner as that in which they are (at least sometimes) defended. But the scientific spirit does not require us to blind ourselves to the practical consequences which hang upon the solution to not a few scientific problems. And assuredly there is no scientific problem upon which so much depends as upon the answer we give to the question whether the distinction which we are accustomed to draw between right and wrong belongs to the region of objective truth like the laws of mathematics and of physical science, or whether it is based upon an actual emotional constitution of individual human beings, which may once have possessed, and possibly may still possess, a certain survival-value in the evolution of the species to which those individual belong. That emotionalist theory of ethics however little intended to have that result by its supporters, is fatal to the deepest spiritual convictions and to the highest spiritual aspirations of the human race* ([15], p. 199–200).

For Rashdall, the problem with intuitionism is not what it tells us about human beings as a product of evolutionary forces beyond their control. Rather, morality is about that part of human evolution that people do control. It sets out ideals, "the highest spiritual aspirations of the human race" in certain terms. What is valuable now is determined on the basis of these ideals, rather than how evolution has shaped us to feel about it. Constructs of human reason, theories and hypotheses, abductions, principles, expressions of "objective truth" like those of mathematical and physical law, tell us what is valuable beyond the range of our evolved capacity to feel about things.

This line of thought represents a strong counter to the nativist push to write reason mostly out of the moral chain of causation. Intuitionists, on Rashdall's account, fail to adequately weigh the consequences of the action. They account for the motivation, the antecedent. But, without objective means to weigh ends of action against one another, when morally salient emotions conflict, it is impossible to decide on the relevant course of action, "for it is impossible to pronounce one motive higher than another in the abstract, without reference to circumstances" ([16], Chap. 4). And, there is no guarantee, or at least not guarantee enough, that evolution has prepared us for the circumstances that confront us at any given moment.[7]

[7] Consider this adaptation of a famously mistaken line, from another famous and oft mistaken utilitarian, John Stuart Mill—The only way that we can know what is worth seeking or avoiding is because these are actively sought or avoided. On Mill's account, you would be better off Socrates suffering then follow the nose of evolution to the end of history. Because, without some power to determine the situation into which life places a human being, that human being remains a slave, thus failing to qualify for moral consideration, at all. Emotions remain stuck in the situation as it is, and insofar as humanity binds its sights to an emotional moral mooring, regardless if these have an evolutionary basis, it is as if mankind had never crawled from the primordial muck, leaving behind its correspondent morality, and adapted to the world as it should be. On the other hand, moral reason attaches to morally ideal situations and principles, because it aims at the best possible consequence regardless of how we feel about it. It is not what evolution has brought us to, but what we do with who we are, today, and tomorrow, these are the ethically

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 20 | Date: 9-5-2013 | Page: 8/28

In this light, consider Peter Singer's position, that the contribution to moral philosophy from the cognitive sciences may be negative, confirming only those aspects of morality that should be pared away in pursuit of adequate moral theories. On Singer's assay, only moral skepticism is the alternative to "the ambitious task of separating those moral judgments that we owe to our evolutionary and cultural history, from those that have a rational basis," with the full intention of discarding all those without a rational basis ([17], p. 351). And, so far as neuroethicists over-confidently swinging the theoretical pendulum are concerned:

> *Advances in our understanding of ethics do not themselves directly imply any normative conclusions, but they undermine some conceptions of doing ethics which themselves have normative conclusions. Those conceptions of ethics tend to be too respectful of our intuitions. Our better understanding of ethics gives us grounds for being less respectful of them* (p. 349).

It is not what evolution has brought us to, but what we do with who we are, today, and tomorrow, these are the ethically relevant aspects of moral life worth talking about. Any evolved, innate emotional dimensionality may describe what we do on the basis of how it feels, but it does not tell us what should be done, regardless, and it is unlikely that revealing the structures further underlying moral reasoning is going to do so, either. Intuitions, insights there into and their theoretical offspring, are merely imperfect starting points to responsible moral agency, and those who hold innate processes as upper and lower limit to the space of moral theory are at best misinformed and at worst naive.

As expressions of our highest, most distinctly human capacities to conceive of ourselves, our world, and the world that we leave behind after actions right or wrong accumulated, these rationalist constructions pull us forward, rather than push us along. They tell us why we live, not just why it feels good or bad when we do it this way or another. And this is their purpose. They open up the space of goal-oriented categories, allowing a currently bad feeling to be endured for a better one. Without these goals, and especially without their development into philosophical ideals, there is no possibility for the analysis of consequences.

(Footnote 7 continued)
relevant aspects of moral life worth talking about. What decides between the emotions is the purpose, the rationally constructed ideal end and object of the action and the emotion that wins is the one that brings about the best possible moral situation consonant with that action's purpose. Any evolved, innate emotional dimensionality may describe what we do on the basis of how it feels, but it does not tell us what should be done, regardless, and it is unlikely that revealing the structures further underlying moral reasoning is going to do so, either.

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 20 | Date: 9-5-2013 | Page: 9/28

3 Two Moral Templates

All of the evidence points to the fact that "Morality is a natural phenomenon. No myths are required to explain its existence" ([17], p. 337). And this clarity extends to all levels of human conduct, with Jonathan Haidt asserting that "Moral systems are interlocking sets of values, practices, institutions, and evolved psychological mechanisms that work together to suppress or regulate selfishness and make social life possible" ([9], p. 70).

The issue is what we do with this understanding, not only to make social life "possible" but to make it better. One way in which this already happens involves tempering immediate desire for long-term cooperative goals. Likewise, Darcia Narvaez warns against the reduction of moral motivation to intuition and emotion due to the limits of "gut-reaction" assessments in both picking out and assigning adequate significance to morally salient features of complex and changing situations. Narvaez points out that morality requires an individual "to step away from his own interests and from current norms to consider more inclusive and logically just forms of cooperation" ([14], p. 167) utilizing all forms of information available in the construction of moral ideals and principles that help us to work together toward more just arrangements.

The ability to create and to set out for one's self moral ideals and ideal situations, better situations, as well as to empathize with others, taking up their situations as if one's own, "in their shoes" so to speak, is moral imagination. Lorenzo Magnani and Emmanuel Bardone characterize moral imagination as "analogical and metaphorical reasoning" that is "very important" to the practice of ethics "because of its capacity to "re-conceptualize" the particular situation at hand," representing the situation as it should be or could have been [18]. Building from work done by Magnani (2001), they suggest that analogical reasoning is a type of model-based reasoning. That being so, moral imagination sets out situations to be sought and others to be avoided, based on information from one's own and others felt, expressed, embodied situations [19].

Building from work done by Magnani, (2007), Magnani and Bardone note another way in which model based reasoning sheds light on moral cognition [20]. Ends set out and achieved may be worked toward without something like what Stich noted earlier in terms of other forms of knowledge, without "conscious access," with agents remaining able to execute sophisticated patterns of behavior, along the "how/that" distinction in epistemology generally. Magnani and Bardone review the notion of "tacit templates" to account for "embodied, implicit patterns of behavior" ([18], p. 100) that are essentially context specific routine actions either non-reflexively triggered through prior training to "be selected from those already stored in the mind–body system, as when a young boy notices his baby sister crying and, without thinking, automatically tries to comfort the infant," or "*created* in order to achieve certain moral outcomes" (authors' emphasis, p. 100). This process of developing a model routine and internalizing it in self-direction,

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 20 | Date: 9-5-2013 | Page: 10/28

toward some further goal, is an illustration of the constructive role of what Magnani has developed as "moral mediators."

Specifically, the sort of model that we are after here is an example of a "task-transforming" external representation. This kind of representation simplifies an otherwise complex task by transforming "difficult tasks into ones that can be done by pattern matching," thereby making possible solutions to problems at hand "transparent," with the understanding that "The more transparent the agent makes the task, the easier it is to find the proper solution" (p. 103).

In this section, we will look to two candidate sources for the sort of task-transforming representations necessary.

First, let's look at Jonathan Haidt's "social intuitionist model." Haidt defines moral intuition as a capacity to realize moral truth without an exercise of reason, but rather through motivating emotions, with the content of these intuitions including emotional valences on the model of perception, with the shape of these valences ultimately due to evolution, recognizing that "it is very difficult to create a fear of flowers, or even of such dangerous things as knives and fire, because evolution did not 'prepare' our minds to learn such associations" ([21], p. 58). Supporting these evolved moral processes are moral modules, "small sets" of which are productive of moral intuitions, and Haidt and Joseph posit the existence of four fundamental sets of modules concerned with purity, reciprocity, hierarchy and suffering. Paying special attention to that concerned with purity, Haidt and Joseph paint a compelling portrait of the extension of moral principle from innate neural structure, providing a universal basis for morality grounding the common forms and functions of moral principles active in different cultures, regardless of apparent differences:

> *Over time, this purity module and its affective output have been elaborated by many cultures into sets of rules, sometimes quite elaborate, regulating a great many bodily functions and practices, including diet and hygiene. Once norms were in place for such practices, violations of those norms produced negative affective flashes, that is, moral intuitions* ([21], p. 60).

The social intuitionist model has "four links" ([22], p. 818). These proceed stepwise as follows. First, the "intuitive judgment link" by way of which "moral judgments appear in consciousness automatically and effortlessly." Second, the "post hoc reasoning link" "in which a person searches for argument that will support an already-made judgment." Third, the "reasoned persuasion link" in which a person communicates his moral reasons to others, and may persuade others by "triggering new affectively valenced intuitions in the listener." Finally, the "social persuasion link" is a passive mechanism potentiated by human sensitivity to "group norms" such that "the mere fact that friends, allies, and acquaintances have made a moral judgment exerts a direct influence on others, even if no reasoned persuasion is used" most notably to agree with allies and friends and to regard others vice versa, resulting in social cohesion through a mechanism not unlike that detailed in Magnani, 2011 [23].

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 20 | Date: 9-5-2013 | Page: 11/28

The "central claim" of Haidt's nativist model is that "moral judgment is caused by quick moral intuitions and is followed (when needed) by slow, post facto moral reasoning" ([22], p. 817). Moral reasoning is defined narrowly as "conscious mental activity that consists of transforming given information about people in order to reach a moral judgment." The social intuitionist model "gives moral reasoning a causal role in moral judgment but only when reasoning runs through other people" because "reasoning is rarely used to question one's own attitudes or beliefs" ([22], p. 819). Haidt defends this hypothesis partly on the basis that challenging comfortable prior evaluations, judgment, or beliefs is resisted due to the fact that these re-evaluations threaten existing self-conceptions and world-views according to which life is interpreted as meaningful and on the right track. This leads reason to the exercise of self-defense, as if a "lawyer" rather than a "scientist," either with different object notions of truth. He also cites memory bias, status-quo and self-interest to motivate a "make-sense epistemology" in which "the goal of thinking is not to reach the most accurate conclusion but to find the first conclusion that hangs together well and that fits with one's important prior beliefs" ([22], p. 819).

Consistent with the "social persuasion link," in which an exemplar, prototype, or demonstration of a moral judgment may lead others to follow suit, Haidt and Joseph assert the superiority of the virtue theoretic approach over other approaches to moral development in that "it sees morality as embodied in the very structure of the self, not merely as one of the activities of the self," with virtues themselves represented as "social skills" "closely connected to the intuitive system," the possession of which are evidenced by "the proper automatic reactions to ethically relevant events and states of affairs"[8] ([21], p. 61). Moreover, as the criteria according to which moral action and moral character are commonly evaluated are virtues relative culture and practice, Haidt and Joseph suggest that we take advantage of the body's "preparedness" to make some associations that expedite learning about those things over others.

But, how to take advantage of this preparedness? Haidt and Joseph differentiate their approach from traditional virtue ethics according to a relative de-emphasis of cultural-environmental determinations of virtue, and increased emphasis on "a smaller number of phenomena that are located more in the organism than in the environment," at once recognizing the central importance of each moral agent's unique embodied situation in the instruction of moral virtue through the inculcation of appropriate "flashes" of moral intuition:

> *These flashes are building blocks that make it easy for children to develop certain virtues and virtue concepts. For example, when we try to teach our children compassion, we commonly use stories about mean people who lack those virtues. While hearing such stories children feel sympathy for the victim and condemnation for the perpetrator. Adults cannot create these flashes out of thin air; they can only put children into situations in which these flashes are likely to happen* ([21], p. 63).

[8] The inverse of which being Thagard's "situational distortions."

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 20 | Date: 9-5-2013 | Page: 12/28

Ultimately, the placement into morally instructive situations, so that innately present moral processes are attuned to salient moral dimensions otherwise lacking in experience, is the limiting factor in the growth and development of moral agency. Of course, Haidt's portrait recalls Rousseau's famous farmer's plot demonstration in *Emil*, and as well suffers from a singular objection. It is not so easy constructing these situations.

Furthermore, picture the eventuality of generation after generation putting selves and children into situations that feel, from a common evolutionary basis, like the right situations to be in. How is this different than the arbitrary hand of nature circling the thread of human fate back on itself? From whence does the hero arise who breaks this cycle and frees the future from the past?

It is the task of the exemplar to demonstrate this sort of information that is impossible to represent otherwise. These sources of moral knowledge do produce those flashes of understanding, while also contributing information about timing, and fine motor action, as well as affective cues signaling appropriate motivations and social cues. However, Socrates is more than two centuries dead. Christ, Mohammed, the great leaders in Martin Luther King, Jr., and Gandhi, all dead and lest we wait for exposure to a possible hero, it is up to us to stand in for absent exemplar. Consider, again that history is a circlet, nose to tail. And, we are back at the beginning, rehashing the same old controversies.

Should we wish to encourage the origination of such moral exemplars, potentiated by moral models, to change the world, is this the most effective way to do it? Where the task is moral self-regulation and philosophical self-determination, a successful moral mediator in the form of a tacit template for moral becoming must simplify this process while at once making solutions to everyday moral problems transparent to the subject. Though such illustrations as Haidt's do render solutions transparent, they do nothing to make them easier to reproduce. This means that they are hard to use, represented in ways that people cannot readily employ. If, indeed, facilitating moral agency is the goal of a moral model, then it is difficult to see how nativist mechanisms can further this goal.

After all, it is impossible to expose a child to the all the necessary right things at the right times, and for many, moral life is mostly a series of corrections on what had been a childhood full of bad information. If we take this notion seriously, and we should, then the direction that Haidt's model is taking us begins shed light on the possible source of a standard for moral worth around the exercise of available agency to re-direct and refine the given moral life.[9] This is an expression of virtue, if it is possible at all.

[9] On Martin Heidegger's account, a person does not choose where and what and who he or she is. Rather, he is "thrown" into a situation, and is left to come to terms with it, to discover it and to understand it, and to courageously become what is necessary to take up his thrown condition, its history, and its people, and employ what potential he can to moving that situation forward, toward a morally ideal situation consonant with an essentially social yet individually embodied condition, all while confronted by its inevitable conclusion in death. A person is governed by moods, with mastery over moods necessary for moral freedom, especially mastery over the dread

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 20 | Date: 9-5-2013 | Page: 13/28

If we allow that it is possible, then this, the "honest toil" of moral agency rests not in availing to hardwired precepts, but rather in moral education, self-development and self-regulation, with the leverage points to affect this process most often outside-in and top-down. That is, it typically requires leisure, self-reflection, and a good bit of luck to borrow from Aristotle. Accordingly, one might object to this briefest of presentations of Haidt's intuitionism on the grounds that self-reflection, the thinking part, is not given due consideration, after all.

Further evidence against intuitionist accounts of morality might also be derived from research proposing that a specific morally motivating emotion does not exist. Batson, for example, locates what others consider moral motivation in selfish gain through "moral hypocrisy," the successful presentation "as-if" being moral without motivation to become so [24]. However, this position stands against some prima facie evidence to the contrary. If moral motivation is limited to selfishness, how does a moral ideal present itself, at all, let alone universally? Is it that becoming a moral exemplar is simply an ultimate realization of hypocrisy, pursued for its presumed social and material benefit, universally realized and sought after? Given the tragic ends having met many memorable moral exemplars across the cultural-historical continuum, and the inspiring pro-social influence their examples continue to have on people around the world, this seems unlikely.

A better answer to these questions may be found in the universal structure and function of moral cognition involving the integration of intuitive and rational mechanisms into the unified prospective concerns of a morally self-regulating entity and fundamental unit of moral value, a structure understood traditionally as conscience.

Space forbids adequate review of the philosophical tradition around conscience. It had been a cornerstone in ethical theory until the late 20th century. Conscience has all but disappeared from moral theory, except for medical ethics where the freedoms of doctors and health care professionals to deliver or to restrict medical attention, care, while constrained by law and business policies that may run contrary to those freedoms, remains a contested issue. In this field, Donald Sulmasy offers a "contemporary" view of conscience deserving brief review here.

Echoing Rashdall's assessment of the importance of our understanding of morality, Sumasy holds that it is "impossible to suggest anything more important to the moral life than conscience." On Sulmasy's account, both individuals and institutions are beholden to conscience, with conscience representing

> *the most fundamental of all moral duties—the duty to unite one's powers of reason, emotion, and will into an integrated moral whole based upon ones most fundamental moral principles and identity* ([25], p. 138).

(Footnote 9 continued)
angst of death. Depending on how far from an ideal situation one is "thrown," more or less work must be done to correct for poor moral upbringing during adulthood in striving toward that morally ideal situation on which his identity rests. See, Heidegger takes Aristotle's "a friend is another me" and makes it a fundamental part of the human condition, *mitdasein*. The other is not another me. The other is me.

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 20 | Date: 9-5-2013 | Page: 14/28

According to Sulmasy, conscience has two aspects, one "turned toward its origin" and the other "turned toward moral acts." It comes to our attention when "deliberating about particular cases." It "establishes a felt need" to act according to "fundamental moral commitment to act with understanding" in a way that maintains moral integrity, by resulting in a situation consistent with personal moral precepts. The established feeling constitutes an evaluative "meta-judgment" over the situations brought about through action, both prospective and retrospective, in the form of guilt or shame associated with unsatisfactory ends, and with peaceful wholeness and integrity the reward for having done the right thing, and having brought about the right end.

Approaching the topic of conscience from the philosophy of psychology and cognitive sciences, Thagard and Finn refer to conscience as "the internal sense of moral goodness or badness of one's own actual or imagined conduct," as a "kind of moral intuition, and as "an indicator of the legitimacy of a moral judgment," bridging innately grounded affect and "internal and external standards" while informing us "about what our moral goals are, as well as about good ways to meet these goals" ([10], pp 150, 168, 161, 161, and 163). This description explicitly unifies "top" and "bottom" processes, with conscience working bottom up in producing what Haidt's model accommodated as emotional valences, this one on the order of rightness and wrongness.

Thagard's model rests in an understanding that "emotions are both cognitive appraisals and somatic perceptions, performed simultaneously by interacting brain areas" (p. 151). Cognitive appraisals are judgments on "the extent to which something aids or hinders our goals." Somatic perceptions are "perceptions of bodily states." Their combination results in a view of cognition that evaluates possible goals in terms of anticipated body states.[10] Conscience, expressed as guilt and shame, thus expresses a situation arrived at in violation of some other emotional valence,[11] and these are not limited to social feelings. Rather, Thagard recognizes the fact that moral and non-moral situations elicit activity in similar regions of the brain, suggesting that there is "nothing special about the brain processes involved in moral intuition compared to emotional consciousness in general." Conscience however, on Thagard and Finn's estimation, concerns moral goals only, such as "increase the well-being of people in general," "act in accord with abstract moral principles such as fairness and respect for autonomy," and "satisfy the expectations of social groups such as family and comply with religious standards or other moral code" (p. 153). This is of course to beg the question—Is it conscience that delineates the moral from non-moral?—but we shall leave this question behind.

In short, conscience, when judging an action right, is expressed as a positive emotional valence associated with the satisfaction of the goal toward which that

[10] Note the parallel with Sulmasy's two dimensional characterization.

[11] We may also deduce that the voice of conscience is anticipated guilt or shame for some situation made possible by some entertained action.

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 20 | Date: 9-5-2013 | Page: 15/28

action aims. Working against these goals results in negative emotions. Thus, conscience represents a mechanism for social compliance, as well as motivations to some other goal for which some positive valence is associated.

When it comes to moral self-regulation and instruction, rather than to conscience, directly, Thagard points to his "informed intuition" model for moral problem solving.[12] This four-step model is decidedly top-down, proceeding thusly:

1. Set up the decision problem carefully. This requires identifying the goals to be accomplished by your decision and specifying the broad range of possible actions that might accomplish those goals.
2. Reflect on the importance of the different goals. Such reflection will be more emotional and intuitive than just putting a numerical weight on them, but should help you to be more aware of what you care about in the current decision situation. Identify goals whose importance may be exaggerated because of emotional distortions.
3. Examine beliefs about the extent to which various actions would facilitate the different goals. Are these beliefs based on good evidence? If not, revise them.
4. Make your intuitive judgment about the best action to perform, monitoring your emotional reaction to different options. Run your decision past other people to see if it seems reasonable to them (p. 162).

This model stands in contradiction to Haidt's hypothesis that "reasoning is rarely used to question one's own attitudes or beliefs." This is a decision procedure seeking reflective equilibrium through a critical evaluation of how given beliefs contribute to the realization of deliberate goals, calling for their revision on this practical basis. Contrary to the intuitionist program, Thagard's takes care to set out ideal situations and evaluate the feelings that arise in their respective consideration, and this gives a critical stance from which to weigh the rationality of given emotional valences. Thus, Thagard's decision procedure goes a long way to answering objections to intuitionism leveled from the likes of Rashdall while remaining sensitive to motivating emotions, and opening the decision process to others who may be affected by actions in question.

But, is this the best way to represent morality in order to further the purpose of moral models, facilitating moral becoming? It certainly stands as an improvement, of sorts, over the virtue approach in that it can be applied in the consideration of hypothetical situations under one's own self direction.

Thagard's decision procedure breaks free from affective limits, and right that it should. Due attention must be given to what constitutes morality in addition to affect, specifically sources of moral freedom rather than evolved pre-determination.[13] The effective and affective detachment from immediate environmental

[12] Which may in moral cases perhaps be called the "educating your conscience" model

[13] After all, if we are not free to determine for ourselves what is right and wrong, and further to act toward one and away from the other, then any talk of morality rapidly reduces to pharmacology.

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 20 | Date: 9-5-2013 | Page: 16/28

pressures is a source of human freedom, with this capacity archetypically realized as syntactical, symbolic, "offline" processing consistent with the perceptual basis of symbols and linguistic representations. (c.f. [26]) In this dimension, Thagard's approach to informing moral intuition is on point. However, it is difficult to identify advantages of Thagard's over other heuristics in framing moral problems, such as decision trees and reflective equilibrium approaches.

It is tedious, requires special time and attention to execute outside of the flow of everyday life, and even if beneficial given leisure, it fails to give direction in how to frame moral problems in a way to best inform moral intuitions. Rather, likely due to the view that there is no special set of morally specific modules in the brain, and no specifically moral processes in cognition, moral problems are approached as any other. In every case, required processing is slow, and so not suitable to directly inform some situations, but rather is best employed in reflection during moments of relative leisure, to recall an opportunity dear to Socrates, in order to rehearse and potentiate "the proper automatic reactions to ethically relevant events and states of affairs."[14] As a result, it ultimately fails to render the process of becoming a moral person transparent.

However, it is clear how Thagard's serves as a compliment to Haidt's approach. Ideally, then, a model intended for moral self-development and instruction would marry the approaches of Thagard and Haidt, while taking advantage of embodied moral processing in a way that facilitates moral becoming through making the process of self-transformation, itself, transparent.

4 The Worm and the Mollusc

> *Although science likes to separate component processes for closer analysis, sometimes this gives the wrong impression—as if one can truly separate the person from the situation, reason from emotion, or intuition from unconscious reason.*
>
> —Darcia Narvaez[15]

First, I think that we can begin to make sense of continuing disagreement over the source and shape of morality through two observations and an image.

In my experience, people view cognition in ways that reflect their own cognitive styles, and cognitive styles are forged by the specific character of, and tempered by the breadth and depth of experience. Philosophers spend a lot of time thinking, while others may spend relatively more of their time doing. As philosophers are most often employed as educators, thus, we find people who spend their

[14] In other words, moral autonomy is to be found in the application of reflective analysis and moral imagination towards the preparation of innate capacities to feel, judge, and act, i.e. in the practice of traditional, especially Socratic, ethics.

[15] [27], p. 185.

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 20 | Date: 9-5-2013 | Page: 17/28

time thinking concerned with communicating moral truths to others who more likely spend their time doing. The more that we reflect on emotions, after all, the sooner they are categorized, and it is easy enough to see how, at least in a man's understanding, this pattern of action might coopt an otherwise elephantine emotional life. But, there is no use in it if the elephant isn't frenzied and restless. There is no sense asking "What it feels like" of an analytic moral philosopher, if the philosopher has never felt it. And, if he has never felt it, then what he has left are his categories and conditions, which is where, I think, we started off with Stich in this paper's introduction.

I further suspect that some disagreement over the nature of morality is due to the subtle abuse of the common conception that human neural processing is dual in nature. Involving

> *two distinct systems through which human beings apprehend reality: System 1 is emotional, affective, intuitive, spontaneous and evolutionary prior; System 2 is rational, analytical, reflective and occurred later in our evolution* ([28], p. 175).

Along with this distinction has arisen a torrent of inquiry into the neural substrates of moral processing that has grown increasingly philosophically sophisticated, and controversy has arisen as these inquiries are framed and results interpreted providing physiological bases for phenomena which had been, traditionally, the domain of moral philosophers. As the theoretical reach of neurology into traditional moral philosophy has deepened, conflict has arisen between theorists who take morality and moral agency as an essentially rational exercise in self-determination, a "System 2" or "top-level" product, and those who take it as a product of evolved processing extended from basic operations maintaining physical integrity in the face of changing environmental pressures, as an essentially affective, "System 1," rather than rational activity. Champions of these respective approaches have contrasted their positions in very strong terms, and this has resulted in controversy. Controversy, moreover, that is not new, and perhaps requires not repeating.

Finally, consider that people have tended to regard the ways in which humans differ from other animals rather than their similarities as the locus of moral value, just as they have for reason and consciousness, categorically defining other animals exempt from moral consideration. It is my suspicion that this sort of reasoning, so "intuitive," has contributed to a misunderstanding of morality that remains implicit in attributions of moral value today. There is more in common between caterpillars and human beings than between human beings and most of the rest of the materials in the universe. Is it possible that some of this common structure is crucial to the moral structure of human beings, as well?

Consider the following story from the life of naturalist Jean Henri Fabre as related by Robert Kirkman, about a type of social caterpillar called a "pine processionary." These caterpillars "venture from the shelter of the nest" at night, in single file lined up without gaps, with "each caterpillar adding a strand of silk to the trail laid down by the leader." One day, Fabre looped this thread back on itself. And, as Kirkman quotes,

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 20 | Date: 9-5-2013 | Page: 18/28

> *The unbroken chain eliminates the leader with his change of direction; and all follow mechanically, as faithful to their circle as are the hands of a watch. The headless file has no liberty left, no will; it has become mere clockwork* ([29], p. 27).

They followed in circles for a week. Such life, for a human being, may not seem worth living. There must be more than that, and it is to the difference between human and caterpillar that people have tended to look, with the implication that morality is not on the model of a caterpillar.

But, why?

Who can say that they have not been in the position of those caterpillars, perhaps once, following friends, associates, lovers on courses that only left them spent, lost, hungry and a week behind?

Consider Haidt's portrayal of the embodied condition as a small stick rider atop a massive elephant, ostensibly the driver but vastly overpowered and at the whim of the beast [30]. This illustration represents a correction on the presumption that people are essentially rational agents, and it does something more. It advises how to most effectively direct one's emotionally grounded life. The trick in directing one's life is to get the rider and the elephant working together.

This model has obvious advantages over, say, dualist models. For instance, there is prima facie reason to take good care of the emotional vehicle that is the body, where for the dualist the body may be more limitation than empowering transport. And, it does capture a sense of what it feels like to be a human being in a humorous way that is easy to accept and employ. However, it does not seem to reflect Sulmasy's profile of a Janus-faced mechanism for moral meta-judgment. And, as for our goal of best representing morality for moral development, what Darcia Narvaez calls "moral self-becoming," it is difficult to see how Haidt's illustration can be of much use.[16]

Let's start out for a better representation by returning to the beginning of the paper, to make something of Stich's use of Barsalou in positing more than one mode of representation at work in moral cognition, with

> *the mental representation of "goal derived" categories, such as things not to eat on a diet… may have a format that is quite different from the mental representation of apple, fruit, or dog.*

The implication is that a good model of morality may need to represent morality in more than one way, corresponding with different mental capacities and modes of operation. This characterization feeds Sulmasy's description of conscience, as well, with moral goals associated with or derived from principled moral conviction and the qualification of other representations falling under these goals colored accordingly. Further, according to Sulmasy, "conscientious" persons may change goals upon "learning certain empirical facts," ([25], p. 144) thereby

[16] And, besides, there is a troubling aspect to Haidt's image. This is that there is a man in the position of reason, and this reveals a tacit association between Haidt's conceptions of humanity and of reason that might be taken to locate moral value in reason.

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 20 | Date: 9-5-2013 | Page: 19/28

educating conscience through a directed search for and exposure to such facts as seen in Thagard's informed intuition model.

Haidt's stick-figure elephant rider also represents two modes of representation at work. But, this picture does nothing to clarify the processes that tie these modes together, not in a pro-moral, or in any other way unless one wishes to carry the metaphor of stick rider further—"Be good to your elephant, and your elephant will be good to you," and so on.

Consider, rather, moral cognition on the model of an emotional inchworm ridden by an information processing bivalve. One, the inchworm, reaches one end of itself forward to feel out possible new situations, while the other end remains rooted in the original. Once felt, the bivalve can open to this information in order to determine what being in that situation would confront it with. On this image, there is no separate human rider struggling atop some furious beast. Rather, we have a single organism of two processes, one reaching forward or back to possible situations, and the other processing available information to compare with the still retained original. The inchworm feels out new situations, while the mollusk comes to terms with them. And the end selected is the one that feels best in the terms reached.

This model of cognition represents the dual nature of cognition in a way that these processes are active in the discovery of the world, in the generation of new experience, and also opens avenues to discussion of virtues, such as courage, versus vices such as recklessness, in a very clear manner. In terms so simple as to invite skepticism, courage requires that one come to terms with the situation that he seeks through action. Without this process, the agent is reckless, and ultimately immoral.

Before detailing this image further, let's examine the model of moral cognition from which it arises, the ACTWith model.

The ACTWith model was originally conceived of as a model of philosophical conscience, informed by Ron Sun's CLARION model of human learning [31]. "ACTWith" stands for "As-if Coming-to-Terms-With," representing a processing framework composed of a four-fold cycle that may be pictured as a sort of intuition-reason pump on the model of the human heart, with the heart being traditionally the embodied locus of conscience. The cycle proceeds as follows:

1) As-if (open) coming-to-terms-with (closed)
2) As-if (open) coming-to-terms-with (open)
3) As-if (closed) coming-to-terms-with (open)
4) As-if (closed) coming-to-terms-with (closed).

Open processes gate information into a process, closed operations process that information, with the open "as- if" operations feeling a situation out, and the open "coming-to-terms-with" operations defining the situation accordingly. So, for instance, in the closed/closed mode, the agent may act on the basis of interred information, returning to the open/closed mode, whereby the newly acquired situation after action is first felt out, and so on through the cycle. Similar processing occurs in active compassion and empathy. Feeling out another's situation is

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 20 | Date: 9-5-2013 | Page: 20/28

facilitated by affective and effective cues which provide comportment information and permit their direct embodiment through mirroring of that embodied condition.

In ACTWith notation, during the o/c stage of processing the agent opens to the situation. In the o/o, the agent feels as if in that situation while opening existing terms of understanding to revision on the basis of this new information, and during the third stage, c/o, the agent updates his existing understanding, not only feeling as if in another situation but understanding this as fact. In this mode, an emotionally laden conception of a candidate situation is generated, and this portrait is compared with the original, with the felt difference between them constituting motivation to seek or to avoid that candidate. During the c/c phase, the agent may act toward that situation, or return to the process of farming for more and better ones.

Different cognitive styles arise through the routine commitment of cognitive resources to the different modes of information processing, with the habitual embodiment of these modes in certain types of situations resulting in the development of different personalities and prejudices. Allocation of resources may be conceived of in terms of clock cycles, electrochemical potentials, or simply as time spent engaged in a certain mode of processing. For example, as the relative evaluation of other situations equally means one's own or another's, an agent may be habitually open to his own possible situations (o/o) while remaining indifferent to those of others (c/c). Regarding feeling out another's situation (o/c–o/o), if one's moral cognitive routine commits ample resources to identifying, recognizing, and personally realizing signs of affective and effective states, then this contributes to a certain cognitive style, including the projection of moral archetypes and the emergence of moral exemplars. In particular, the habitual exercise of the o/c/- o/o modes in morally significant situations potentiates exemplary kindness as well as wisdom, due to the fact that experiential resources are rapidly expanded, and bases for common understanding and terms for communication expanded, all contributing to a decidedly pro-social personality type. This cognitive style is "conscientiousness."

The ACTWith model makes easy sense of other basic moral attitudes, too. In compliment to Stich's "Platonic assumptions," consider the following "Socratic precepts" that arise from normal ACTWith operation.

The first of these assumptions is "Know nothing." Socrates was famous for suggesting that, though confirmed the 'wisest man in Athens,' he knew nothing. His method in discovery through discourse involved always beginning with the situation as understood by his interlocutors, and proceeding from there towards an adequate assay of the matter at hand. On the ACTWith model, this is represented by the first steps. In meeting with others, Socrates opens to the situation, then opens to the terms to which they have come in determining the situation, only feeling out and assessing further possible situations after this preliminary stage. By this precept, thus, one must adopt a situation as if one's own in order to begin to know why it is or is not satisfactory, why movement from this position (literal and figurative) is necessary, in order to lead from there to something better. Prior experience is active beginning in the third stage if this Socratic method is modeled

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 20 | Date: 9-5-2013 | Page: 21/28

after, but starting open to "what it feels like" to be in other situations, and informing one's understanding on this basis without prejudice is key. Making this movement habitual is the first step in becoming a conscientious moral agent.

A second Socratic precept is "Never cross your daemon." Socrates was famous for refusing to aid in the arrest and eventual execution of Leon of Salamis, and also for saying that he was gifted with an innate sense of justice, a "daemon" that forbid him from doing the wrong things. All that he had to do, he told us, was not to cross his daemon in order to emerge the 'most just man in Athens.' This function of conscience is represented in the ACTWith model as follows. As the cycle of processing completes, with terms of understanding come to insofar as resources had been dedicated to their assay during the first stages, the c/c stage draws the agent in on itself in preparation for action. Here, the infamous "voice of conscience" may arise, barring action and so barring passage to associated situations. Here, the last and the future situations are held together, at once, by either end of the illustrative inchworm, at the moment that the inchworm may commit, lifting its tail from its prior situation to pull itself forward into the next. Anticipating that chosen end, updating information until the commitment to the new situation is enacted, conscience reveals that progress to this new situation will result in a loss of progress toward some internalized moral ideal self-representation. That is, one feels as if he will no longer be his own best example of life worth living because the agency that results in said situation is contrary to the sense of agency exemplified in one's "highest spiritual aspirations," to become the best person one can possibly become. In this final instant before action, with both situations bridged and the embodiment of the new situation imminent, the agent is confronted by what Kant would call "self-repugnance" or self-disgust at the self that results from this situation. Thus, it is not the end, or the action itself, that are rejected in the "veto" of conscience, but rather what is rejected is the self that one will become through said action and at said end. This characterization captures the way in which conscience associates with integrity, feeling of "wholeness" and self-esteem, in natural, easy to employ terms.

The preceding Socratic precepts represent a traditional understanding of conscience while presenting this understanding in a way that is both consistent with what is understood about the neurology of moral cognition and that takes advantage of what is known about these processes in order to facilitate the self-direction of these processes towards a unifying purpose, moral self-development. These emerge from normal exercise of the ACTWith model. The ACTWith model, moreover, is able to accommodate different accounts of moral cognition, as well, even those that seem contrary to the model itself. These other accounts of moral cognition can be informatively mapped onto the ACTwith operations, showing that the ACTWith model is more fundamental.

Consider the following passage from Adam Smith's *Theory of Moral Sentiments* as he describes the process whereby he comes to understand the moral significance of another's embodied condition. Standard ACTWith notation has been added:

By the imagination we place ourselves in his situation [*O/C*]*, we conceive ourselves enduring all the same torments* [*O/O*]*, we enter as it were into his body*

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 20 | Date: 9-5-2013 | Page: 22/28

[*O/O*], *and become in some measure the same person with him* [*O/O*], *and thence form some idea of his sensations* [*C/O*], *and even feel something which, though weaker in degree, is not altogether unlike them* [*C/O*]. *His agonies, when they are thus brought home to ourselves* [*C/O*], *when we have thus adopted and made them our own* [*C/C*], *begin at last to affect us, and we then tremble and shudder at the thought of what he feels* [*O/C - > C/C, in reflection*] ([32], Sect. 1.1.2).

Similarly, Thagard's guide for informed intuition can also be mapped onto the ACTWith model. And, though Thagard's is not primarily a model of moral cognition, in so far as it is applicable to moral direction it should proceed according to the ACTWith logic if the ACTWith model is successful in articulating a universal structure for moral information processing according to which other approaches can be relatively evaluated and recommended. ACTWith notation and brief interpretive comments are added, as follows:

1. Set up the decision problem carefully. [O/C]—feel out the space of possibility.
2. Reflect on the importance of the different goals. [O/O]—attune one's self to the likely realization of different possibilities.
3. Examine beliefs about the extent to which various actions would facilitate the different goals. [C/O]—refine preconceptions based on expected outcomes.
4. Make your intuitive judgment about the best action to perform, monitoring your emotional reaction to different options. [C/C]—act towards a new situation, then/or repeat the cycle.

And, we can do the same thing with Haidt's four-step social intuitionist model, too:

1. The "intuitive judgment link" by way of which "moral judgments appear in consciousness automatically and effortlessly" is O/C wherein arise gut-reactions to possible situations.
2. The "post hoc reasoning link" "in which a person searches for argument that will support an already-made judgment" is C/O, as terms of understanding are farmed for confirmation of the gut-reaction product of step 1. Note that Haidt effectively skips the O/O step, wherein new terms of understanding are generated, bottom-up, so the C/O stage is rather anemic on Haidt's model, thereby limiting moral development consistent with his presumption that reasons is not part of the chain of moral causation.
3. The "reasoned persuasion link" in which a person communicates his moral reasons to others, and may persuade others by "triggering new affectively valenced intuitions in the listener" is C/C, as the persons perform communicative acts, effectively changing the social dimensions of the situation. Presumably, then, the person will enter into a new cycle of processing from this altered situation, until action toward the realization of the felt goal is potentiated.
4. Finally, the "social persuasion link" representing the "direct influence on others" that morally salient action exerts, "even if no reasoned persuasion is used" seems to be a complex of O/C (open to the demonstrated examples of others), O/O (being directly influenced to follow or to reject those examples),

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 20 | Date: 9-5-2013 | Page: 23/28

C/O (exemplars represent a mode of understanding, with this understanding applied to like situations), and C/C (actively exemplifying virtue or vice as information for others).

It is not troubling that these processes are not replicas or duplicates of the ACTWith model, as they each express different assays of moral cognition consistent with the cognitive styles of their creators. It is merely a sign that the ACTWith model is more fundamentally sound than these others in that the ACTWith model had been designed in order to be able to accommodate these variants, as well as more radical variants such as those demonstrated by psychopaths, both individual and institutional, as well as artificial moral agents, and examples from traditional moral philosophy [33–37].

Some comparisons are in order. There is nothing essentially moral about Thagard's model for informed intuition. Neither is there anything essentially moral about Haidt's "nativist" model. One is an extension of individual prudence endorsed through friendly confirmation in the final step. The other is an extension of primal mechanisms aiming at contextually various satisficing conditions, with moral excellence arising through some unspecified mechanism (though perhaps in the ACTWith spirit due to the projected emotional fit of the organism to some projected ideal moral situation). On the other hand, the ACTWith model is essentially a model of morality. On its account, cognition essentially sets out and weighs potentially embodied situations, not simply one's own and not neglecting that potentials can approach zero. This is all undertaken in energetic terms which, due to common physiology and natural law, provide a universal basis for the relative evaluation of embodied situations, and so provide a universal basis for the moral judgment over any given situation and the actions, conventions, and institutions that bring it about.

Space forbids further details, but, very quickly, perhaps the greatest upshots to this model are the following.

One, it encourages the development of moral exemplars, helping to draw human moral development forward. And, it does this while making consistent sense of ongoing research in moral cognition. For example, the ACTWith model makes sense of recent research that persons who are generally or easily disgusted exhibit harsher moral judgment than others less sensitive to disgust, and that these results can be reproduced when the evaluative basis in mood is temporarily induced through disgusting and irritating noises.

Two, the ACTWith model naturalizes intention in an intuitive and useful way. With conscience understood as the felt comparison of relatively well-ordered situations with the ideally ordered case understood as an ideal arrangement of objects on minimal dimensions, a "-science," and with the felt tension between situations motivational, intention can be understood as "in-tension." Given the common energetic basis of the ongoing analysis of situations on the ACTWith model, intension is understood as the internal, motivating and relatively evaluative felt strain, or "tension," between conscientiously compared situations, reference to which expresses both the motivation to some end as well as the end, itself. This

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 20 | Date: 9-5-2013 | Page: 24/28

interpretation falls in well with everyday language. For example, one "intends" to bring a situation about simply because it is a better situation to be in according to the terms of evaluation brought to bear in the comparison, noting that these terms need not be subject to conscious selection.

One may object that this makes no sense of intentions over individual objects. I think that such a possible objection is mistaken for two reasons. One, there is no compelling evidence that cognition attends to individual objects rather than possible effects that these objects may have on possible situations. One need only consider how dramatically a situation can change when it includes a door key, or a restroom, to see that, as individual objects in the placements and properties change, so do the situations in which they take part. And, moreover confirmed in intuition, the only sense in which these objects do take place, or not, is that in which the situation as a whole is transformed by their presence or lack thereof.

Another upshot for the ACTWith model is that the ACTWith program naturalizes freewill as the embodied metabolic potential to posit, alter, construct and to otherwise act toward ends of one's own self-determination, not least through attending to and altering the weights attached to salient terms brought to bear in rational analysis. Most importantly, this process underwrites philosophical self-determination, the particular capacity of directed thought to affect the sort of person that one will become through action by inculcating automatic or practiced reactions to specific opportunities when so presented. Ultimately, this capacity is due to the fact that thinking about one situation rather than another, in one set of terms rather than another, expends similar amounts of physiological potential, leveling the decision space given relative lack of urgency. Though fundamental to Thagard's informed intuition model, this aspect of moral agency is discounted on Haidt's, but only in the ACTWith model is the metabolic basis for cognition as well as bodily actions rendered in one coherent frame.

Finally, the ACTWith model helps to make sense of otherwise troubling concepts from the philosophical tradition concerning moral self-development, encouraging the aspiration to moral ideals rather than wrote internalization of moral principle or affect, and this deserves the briefest of accounts. By the ACTWith program, conscience signifies the enveloping framework of cognition, guiding an agent from situation to situation. It lays out possible ends of action as situations in which the agent innately seeks to retain integrity by maintaining equilibrium between internal and external forces, and this embodied logic, along with embodied limitations, allows for their comparison and relative evaluation, with differences providing motivation to move toward some and away from others. Fundamental terms for the relative evaluation of situations are derived from metabolic, physiological constraints, and are thus essentially energetic rather than material. Conscience so conceived is the felt comparison of situations in the constant adjustment of any dynamic agent to its changing internal and external environments, in the human instance via homeostatic regulation of embodied processes extending through moral cognition, including the comparison of possible situations hypothesized in terms with which the person already cognizes and acts as made available through limiting experience, i.e. "moral imagination."

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 20 | Date: 9-5-2013 | Page: 25/28

As the constitution of these hypotheticals proceeds from a limited sphere of individual experience, augmented by affective and effective mirroring as well as taught "top-down," there is great potential for the scope of conscience to expand over the course of operation. As terms increase, given sufficient resources, the agent may develop capacities to simultaneously evaluate greater numbers of dimensions and to more readily identify morally salient dimensions. With the space of action mapped through this operation properly understood as meta-physical, rather than merely physical, conscience motivates the agent to seek situations with minimal strain between one's own and others' current and expected future situations, with the global minimum—informed as described, through habitual conscientiousness—specified as the Kantian "summum bonum" [35].

This inspirational quality is obvious from the ACTWith structure. According to Kant, an agent would be merely "a marionette or automaton" without the tension between the sensible and the ideal, with any sense of freedom a "mere delusion," freedom "only in a comparative sense, since, although the proximate determining causes are internal, yet the last and highest is found in a foreign land" ([38], p. 102). Substitute "pine processionary" for "marionette" and the relationship becomes clearer. After all, should a marionette live, it is not a life worth living, perhaps even less so than the caterpillar's and for similar reasons. The source of the motivating moral tension ultimately drawing the moral agent on to the Kantian "kingdom of ends," aspiring to Kantian reverence and away from moral repugnance, is conscience as understood on the ACTWith model.

5 Conclusion

I want to close by reconsidering Rashdall's phrase, introduced earlier, that "the scientific spirit does not require us to blind ourselves to the practical consequences which hang upon the solution to not a few scientific problems."

How we conceive of morality has practical consequences. These conceptions leave morality more or less available to practice. So, conceptions that make solutions to moral problems transparent are the best.

Perhaps the most important moral problem confronting every moral agent is who he will become through a life of action, a good person or bad. The ACTWith model helps to make solutions to this ongoing problem transparent. Moreover, potentiating moral self-determination raises the bar of human leadership, and this is promising for the future of human tolerance and liberty, qualities sadly failing to tyranny in the current era. After all, who willingly serves a lesser man than himself, to lesser ends than he is able, but a slave, or a worm, or a marionette, all without moral significance? This answer to this question is also rendered transparent on the ACTWith model.

"Ultimately, a genuine leader is not a succor for consensus but a mold of consensus" [39]. Leaders do more than make and break laws. They exemplify ways of life, ways which, due to the nature and namesake of their positions, others

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 20 | Date: 9-5-2013 | Page: 26/28

follow, a fact of the human condition to which Haidt gives due attention. Towards these, and for example in "conscientious objection" radically different ends, the ACTWith model facilitates life-long moral development in a practical, holistic way, being an intuitive, quick and transparent heuristic, which, easily employed routinely and habitually entrains the agent into a specifically moral virtue, conscientiousness. In short, where other models ask if an act is prudent, or safe, if it feels good or even if it is popular, the ACTWith model of conscience asks of the proposed end of action, "Is it right?"

For John Dewey, the capacity to imagine other situations, to manipulate those situations, and to relatively weigh them, as is required in assessing the consequences of actions, "constitutes an extension of the environment to which we respond" ([40], p. 387) Imagination confronts the thinker with possible situations, by placing the thinker in those situations, forcing the thinker to come to terms with those situations as if they were his own.[17] This is because cognition is not separate from the body and from its situation. Rather, in Dewey's words, "mind is a complex function of the doings and under goings of encultured, embodied, historically situated organisms, continuous with physical systems" ([41], p. 10).

This understanding, nearly a century old, is worthy of claiming today. And this reveals something about the tradition in moral philosophy and the future of moral theory. Though the cognitive sciences have contributed to our deepening understanding of the wheels that turn within us, it has offered less in the way of self-regulatory powers over those same processes. Intuitions and their evolutionary origins do not directly show us how to succeed in becoming moral, to remain so, or to aspire to some higher level of moral virtue. Moreover, such a neurologically based understanding of morality is not easily applied in the evaluation and similar reform of institutions and collectives, themselves by some regarded as morally significant individuals in their own right. As well, neurological models are useful, but by no means prescriptive in considerations of the engineering design and moral standing of artificial moral agents, or any other morally significant entity, individual or collective, so far beyond study. The ACTWith model of moral cognition was developed to overcome these shortcomings.

In the end, our deepening understanding of embodied moral mechanisms may not be the most important tool in our moral development. And this returns us to the inspiration that set us out on this journey, Stich's call to collaboration on the most important questions in moral life. With Stich, in the beginning of this paper, we found moral philosophy chasing its own tail, without the influence and information from other disciplines, especially psychology. Here, at the end of our discussion, do we not find the cognitive sciences chasing its own tail? After all, in testing for morally salient functionality specific to certain areas of anatomy, do the scientists not test from the same set of action potentials and expectations that guide their

[17] This portrait is supported by evidence that similar pathways of neural processing "are activated both during prospection and during hypothetical moral decision-making," ([40], p. 749) and that all cognition is essentially of the embodied condition.

own subjective experience? They confirm, then, only themselves in what they study. Their work reflects their evaluations, and expectations, as these are all that they know to challenge. But, what of moral ideals? Where are these to be tested, weighed, measured? Is it not from philosophy, and not cognitive science, that any question as to the potential realization of this human body arises? And without this view to the human future, what is the value of anything, at all, but what it is rather than what it might become?

With these questions in mind, let's close with some reflections on the future on moral philosophy from Young and Koenigs. Though they show no doubt that extra-rational processes play decisive roles in moral judgment, for better or for worse, given that "A coarse summation of the clinical findings is that individuals who exhibit abnormal emotional processing also exhibit systematically abnormal moral judgment," these scientists note that, perhaps, the pendulum of progress into the question of moral representations has reached its zenith in the cognitive sciences. They tell us that "Even though the acquisition or expression of moral knowledge may be a suitable subject of scientific inquiry, science cannot reveal what is morally right or morally wrong," and that the "brain may thus constrain the moral mind, but how we decide to deal with such constraints may be best determined in philosophical debate." Finally, looking forward, they point back to moral philosophy, and back in the direction from which we have come. Their advice is to "return to the likes of Kant, Hume and Mill or join the efforts of a new camp of scholars, empirical philosophers, who seek to marry descriptive and normative approaches to human moral psychology" ([42], p. 77). Advice worth following.

References

1. Schopenhauer, A.: The Basis of Morality. (Translated with introduction and notes by A.B. Bullock.) Swan Sonnenschein & Co., London (1903)
2. Stich, S.: Moral philosophy and moral representation. In: Hechter, M., Nadel, L., Michod, R. (eds.) The Origin of Values. Aldine de Gruyer, New York (1993). http://www.unc.edu/~knobe/x-phi/stich.pdf
3. Andersen, N.: Conscience, recognition, and the irreducibility of difference in Hegel's conception of spirit. Ideal. Stud. **35**(2), 119–136 (2005)
4. Eisenberg, N.: Emotion, regulation, and moral development. Annu. Rev. Psychol. **51**, 665–697 (2000). http://psych.colorado.edu/~tito/sp03/7536/eisenberg_2000.pdf
5. LeDoux, J.: Rethinking the emotional brain. Neuron **73**, 653–676 (2012)
6. Osman, M.: An evaluation of dual-process theories of reasoning. Psychon. Bull. Rev. **11**, 988–1010 (2004)
7. Haidt, J.: The new synthesis in moral psychology. Science **316**, 998–1002 (2007)
8. Kauppinen, A.: Intuition and belief in moral motivation. In: Björnsson, G. (ed.) Moral Motivation: Evidence and Relevance. Oxford Univ. Press, Oxford (in press). http://tcd.academia.edu/AnttiKauppinen/Papers
9. Haidt, J.: Morality. Perspect. Psychol. Sci. **3**, 65–72 (2008)
10. Thagard, P., Finn, T.: Conscience: what is moral intuition? In: Bagnoli, C. (ed.) Morality and the Emotions, pp.150–169. Oxford University Press, Oxford (2011)

11. Krause, J.: Collective intentionality and the (re)production of social norms: the scope for a critical social science. Philos. Soc. Sci. **42**, 323–355 (2012)
12. Young, L., Saxe, R.: Moral universals and individual differences. Emot. Rev. **3**(3), 323–324 (2011)
13. Cokely, E.T., Feltz, A.: Adaptive variation in judgment and philosophical intuition. Conscious. Cogn. **18**, 356–358 (2009)
14. Narvaez, D.: Moral complexity: the fatal attraction of truthiness and the importance of mature moral functioning. Perspect. Psychol. Sci. **5**, 163–181 (2010)
15. Rashdall, H.: Is Conscience an Emotion? Three Lectures on Recent Ethical Theories. Houghton Mifflin, Boston (1914)
16. Rashdall, H.: The Theory of Good and Evil: A Treatise on Moral Philosophy. Oxford University Press, London (1924)
17. Singer, P.: Ethics and Intuitions. J. Ethics **9**, 331–352 (2005)
18. Magnani, L., Bardone, E.: Distributed morality: externalizing ethical knowledge in technological artifacts. Found. Sci. **13**(1), 99–108 (2008)
19. Magnani, L.: Abduction, Reason, and Science. Processes of Discovery and Explanation. Kluwer Academic/Plenum Publishers, New York (2001)
20. Magnani, L.: Semiotic brains and artificial minds. How brains make up material cognitive systems. In: Gudwin, R., Queiroz, J. (eds.) Semiotics and Intelligent Systems Development. Idea Group Inc., Hershey (2007)
21. Haidt, J., Joseph, C.: Intuitive ethics: how innately prepared intuitions generate culturally variable virtues. Daedalus **133**, 55–66 (2004)
22. Haidt, J.: The emotional dog and its rational tail: a social intuitionist approach to moral judgment. Psychol. Rev. **108**(4), 814 (2001)
23. Magnani, L.: Understanding Violence. Springer, Dordrecht (2011)
24. Batson, C.D.: What's wrong with morality? Emot. Rev. **3**, 230–236 (2011)
25. Sulmasy, D.: What is conscience and why is respect for it so important? Theor. Med. Bioeth. **29**, 135–149 (2008)
26. Barsalou, L.W.: Perceptual symbol systems. Behav. Brain Sci. **22**, 577–660 (1999)
27. Narvaez, D.: The embodied dynamism of moral becoming: reply to Haidt. Perspect. Psychol. Sci. **5**, 185–186 (2010)
28. Roeser, S.: Intuitions, emotions and gut reactions in decisions about risks: towards a different interpretation of 'neuroethics'. J. Risk Res. **13**, 175–190 (2010)
29. Kirkman, R.: Through the looking-glass: environmentalism and the problem of freedom. J. Value Inq. **36**(1), 29–43 (2002)
30. Haidt, J.: The Happiness Hypothesis: Finding Modern Truth in Ancient Wisdom. Basic Books, New York (2006)
31. Sun, R.: Duality of the Mind: A bottom-up approach toward cognition. Mahwah, N.J: L. Erlbaum Associates (2001)
32. Smith, A.: The theory of moral sentiments: Raphael, D.D., Macfie,A.L. (eds.) Glasgow Edition of the Works and Correspondence of Adam Smith, vol. I. Liberty Fund, Indianapolis (1982). http://oll.libertyfund.org/title/192
33. White, J.: Manufacturing morality, a general theory of moral agency grounding computational implementations: the ACTWith model. In: Floares, A. (ed.) Computational Intelligence. Nova Science Publishers, Hauppauge (2012)
34. White, J.: An information processing model of psychopathy and anti-social personality disorders integrating neural and psychological accounts towards the assay of social implications of psychopathic agents. In: Fruili, A.S., Veneto, L.D. (eds.) Psychology of Morality. Nova Science Publishers, Hauppauge (2012)
35. White, J.: Autonomy rebuilt: rethinking traditional ethics towards a comprehensive account of autonomous moral agency. Nat. Intell. **1**, 32–39 (2012)
36. White, J.: Conscience: toward the mechanism of morality. University of Missouri-Columbia (2006)

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 20 | Date: 9-5-2013 | Page: 29/28

37. White, J.: Understanding and augmenting human morality, the ACTWith model. In: Magnani, L, Pizzi, C., Carnielli W. (eds.) Studies in Computational Intelligence #314: Model-Based Reasoning in Science and Technology, pp. 607–620. Springer, Heidelberg (2010)
38. Kant, I.: The Critique of Practical Reason, (trans. Abbott, T.K. 1788) Pennsylvania State University Electronic Classics Series (2010). http://www2.hn.psu.edu/faculty/jmanis/kant/Critique-Practical-Reason.pdf
39. King, M.L., Jr.: The other America. http://www.gphistorical.org/mlk/mlkspeech/index.htm (1968)
40. Alexander, T.: John Dewey and the moral imagination: beyond Putnam and Rorty toward a postmodern ethics. Trans. Charles S. Peirce Soc. **29**, 369–400 (1993)
41. Fesmire, S.: John Dewey and moral imagination: pragmatism in ethics. Indiana University Press, Bloomington (2003)
42. Young, L., Koenigs, M.: Investigating emotion in moral cognition: a review of evidence from functional neuroimaging and neuropsychology. Br. Med. Bull. **84**, 69–79 (2007)